

't Hooft suppression and holographic entropy

William R. Kelly, Kevin Kuns and Donald Marolf

*University of California, Santa Barbara,
Santa Barbara, CA 93106, U.S.A.*

E-mail: wkelly@physics.ucsb.edu, kuns@physics.ucsb.edu,
marolf@physics.ucsb.edu

ABSTRACT: Recent works have related the bulk first law of black hole mechanics to the first law of entanglement in a dual CFT. These are first order relations, and receive corrections for finite changes. In particular, the latter is naively expected to be accurate only for small changes in the quantum state. But when Newton's constant is small relative to the AdS scale, the former holds to good approximation even for classical perturbations that contain many quanta. This suggests that — for appropriate states — corrections to the first law of entanglement are suppressed by powers of N in CFTs whose correlators satisfy 't Hooft large- N power counting. We take first steps toward verifying that this is so by studying the large- N structure of the entropy of spatial regions for a class of CFT states motivated by those created from the vacuum by acting with real-time single-trace sources. We show that $1/N$ counting matches bulk predictions, though we require the effect of the source on the modular hamiltonian to be non-singular. The magnitude of our sources is ϵN with ϵ fixed-but-small as $N \rightarrow \infty$. Our results also provide a perturbative derivation — without relying on the replica trick — of the subleading Faulkner-Lewkowycz-Maldacena correction to the Ryu-Takayagi and Hubeny-Rangamani-Takayanagi conjectures at all orders in $1/N$.

KEYWORDS: Gauge-gravity correspondence, AdS-CFT Correspondence, $1/N$ Expansion

ARXIV EPRINT: [1507.03654](https://arxiv.org/abs/1507.03654)

Contents

1	Introduction	1
2	Setting and assumptions	3
2.1	Adapting the source	5
3	Results with no singular terms	7
3.1	Energy	7
3.2	Entropy	8
3.3	Multi-trace contributions to ΔS_A	9
3.3.1	Reorganizing the action of single-trace sources	9
3.3.2	The entropy as a correlator	11
3.3.3	Counting powers of N	13
4	Discussion	14
A	Computing the relative entropy	17
A.1	The Baker-Campbell-Hausdorff formula	18
A.2	Computing $\delta^2 R_A$	18

1 Introduction

There has been much activity exploring the intriguing connection between entanglement in holographic field theories and the gravitational field equations of the bulk dual. This program traces its roots to Jacobson’s seminal paper [1], which proposed the Einstein equation to be a thermodynamic equation of state for some unknown quantum mechanical system in which the area of surfaces measures entanglement entropy across causal horizons. Several groups [2–5] have now used related arguments to derive the linearized gravitational field equations in the context of the anti-de Sitter/conformal field theory (AdS/CFT) correspondence, where the underlying quantum system is well understood; see also the related works [6–8]. A key assumption in these more recent derivations is the Ryu-Takayangi (RT) conjecture [9, 10] and its covariant generalization by Hubeny-Rangamani-Takayanagi (HRT) [11], both of which relate entanglement entropy of subregions of the field theory to the geometry of bulk surfaces. A partial converse of this result — that the bulk field equations imply aspects of the Ryu-Takayanagi conjecture — has also been argued by Lewkowycz and Maldacena [12], though see [13] for a discussion of the so-called homology constraint, [14] for questions about the possible role of complex bulk surfaces, and [15] for concerns regarding strong time-dependence.

Much of the recent discussion has centered on the so-called first law of entanglement

$$\delta S_A = \delta \langle H_A \rangle. \quad (1.1)$$

Here A is a subregion of some Cauchy surface for the CFT, $S_A := -\text{Tr}(\rho_A \log(\rho_A))$ is the von Neumann entropy of the associated reduced density matrix ρ_A , $H_A := -\log(\rho_A)$ is the modular Hamiltonian, and δ denotes the first variation with respect to the state when the operator H_A is held fixed on the right hand side. The relation (1.1) holds at first order, but receives corrections for finite changes.

According to the HRT conjecture, in holographic theories the left hand side of (1.1) can be extracted from the bulk geometry. The right hand side is typically difficult to evaluate, though it reduces to a simple integral of the CFT stress tensor [16] (see also [17]) when A is a ball-shaped region and the system is in its global vacuum state. The combination of these two results makes (1.1) a useful formula for studying the relationship between entanglement and geometry. In particular, in this context (1.1) coincides with the first law of black hole mechanics in the bulk applied to the Rindler-like Killing horizon defined by a ball-shaped region A on the AdS boundary [17]. This argument for the first law can then be inverted to derive linearized bulk equations of motion from HRT [2–5].

The starting point for our work is the observation (see e.g. [4]) that the entanglement first law (1.1) is generally useful only for infinitesimal changes in the quantum state; higher order corrections tend to make significant contributions when it instead undergoes any substantial change. In contrast, the bulk first law of black hole mechanics accurately describes classical deformations — typically involving very large numbers of quanta — so long as the changes in entropy and energy are small in comparison with their background values. In particular, corrections to the bulk first law stem from gravitational back-reaction and are thus suppressed by powers of the bulk Newton constant G_N . This suggests that corrections to the entanglement first law (1.1) will be correspondingly suppressed, at least for states that would be appropriately semi-classical with respect to the bulk. Our goal below is to show suppressions by powers of N in similar computations involving the entropy of spatial regions for CFTs whose correlators satisfy ’t Hooft power counting and whose spectrum of light operators is sufficiently sparse. We consider a class of CFT states motivated by those created from the vacuum by acting with real-time single-trace sources, though we require the source to have non-singular effects on the modular Hamiltonian. In order to model sources that would produce semi-classical coherent states in any bulk dual, the magnitude of our sources is taken to be ϵN with ϵ fixed-but-small as $N \rightarrow \infty$. See also [18] for other investigations of entanglement in large N gauge theories.

For such states, we verify bulk predictions of powers of N via a direct calculation in the CFT. Though it differs in detail, the suppression found here is analogous to the large N suppression of such corrections found previously in [19, 20]. Our analysis also provide additional benefits. First, our explicit formula for second order relative entropy makes manifest the agreement with appropriately-integrated bulk stress tensors — here defined to include contributions from the stress tensor of bulk gravitons — required by comparing corrections to the bulk and boundary first laws [21, 22]. As a result, it again demonstrates that a consistent holographic theory of gravity must couple universally to

all forms of bulk stress-energy. It also provides a perturbative derivation of the Faulkner-Lewkowycz-Maldacena subleading correction to the RT and HRT conjectures at all orders in $1/N$ [23].

Before proceeding, we should elaborate on the above restriction to sources with non-singular action on the modular Hamiltonian. As explained in section 2 below, from a dual bulk point of view this requires our perturbations to vanish in some neighborhood of the bifurcation surface of our bulk Rindler-like horizon. So the leading-order large- N RT or HRT entanglement cannot change. They do, however, affect the above-mentioned order N^0 entanglement. They also change $\langle H_A \rangle$ and thus the relative entropy $R_A = \langle H_A \rangle - S_A$ at order N^2 . We will show that these powers of N are correct at all orders in ϵ ; the fact that ΔS_A is of smaller order in N than S_A itself is the $1/N$ suppression advertised above.

The rest of the paper is organized as follows. We begin with a brief description of our setup in section 2. Section 3 then computes $\langle H_A \rangle$, R_A and S_A to second order for our family of states. It also argues to all orders that the powers of N in $\langle H_A \rangle$, R_A and S_A are precisely those predicted by intuition from a bulk dual. We conclude in section 4 with a brief discussion of our results as well as comments on possible extensions. Appendix A derives a simple and completely general formula (A.14) for the second order change in R_A in bipartite quantum systems that gives the above-mentioned explicit formulas for R_A and ΔS_A .

2 Setting and assumptions

We wish to study excitations of the vacuum $|0\rangle$ of a large N CFT in d spacetime dimensions on $\mathbb{R} \times S^{d-1}$. Since we take our inspiration from a possible bulk dual, we impose assumptions similar to those in e.g. [24] and [25, 26], taking our CFT to satisfy 't Hooft large- N factorization [27] and to have a sparse spectrum of light operators. As usual, light operators are those whose scaling dimension Δ_i of $\mathcal{O}^i(x)$ remains finite as we take $N \rightarrow \infty$ and the sparse spectrum condition requires that for any fixed Δ the number of such operators with $\Delta_i < \Delta$ remains finite at large N .

The factorization condition states that the set of light single-trace gauge-invariant local operators should admit a basis $\{\mathcal{O}^i(x)\}$ for which

$$\langle \mathcal{O}^{i_1}(x_1) \dots \mathcal{O}^{i_k}(x_k) \rangle_c \sim N^{2-k}, \quad (2.1)$$

where $\langle \dots \rangle_c$ is the connected vacuum correlator and similar notation without the subscript c will also be employed for the full correlator. Furthermore, heavy operators (those that do not remain light as $N \rightarrow \infty$) decouple in the sense that connected correlators involving both heavy and light operators are much smaller. In fact, we assume that we study a low energy process from which operators of finite-but-large dimension are sufficiently decoupled that sums over the \mathcal{O}^i below always converge. Indeed, our only use of the sparse spectrum condition will be to assume that such sums with coefficients of order N^p converge to a result of the same order in N .

It will be convenient to take the basis operators \mathcal{O}^i to have definite scaling dimension Δ_i at large N and in fact to diagonalize the order N^0 term in the connected two-point

function. We also require

$$\langle \mathcal{O}^i(x) \rangle = 0, \quad (2.2)$$

for all i , which can be achieved by subtracting appropriate expectation values. Note that we have not required the \mathcal{O}^i to be scalars; we have merely suppressed any tensor or spinor indices. Thus, up to the above subtractions, one member (say, \mathcal{O}^0) of our basis is $1/N$ times the CFT stress tensor which necessarily satisfies $\Delta = d$; i.e. $\mathcal{O}^0 = \frac{1}{N}(T - \langle T \rangle)$, where we have suppressed the spacetime indices ab .

Our expectations that corrections to the first law of entanglement are suppressed arise from considering the semi-classical behavior of a supposed bulk gravitational dual. Semi-classical bulk states can be created from the vacuum through the action of large classical sources for the perturbative bulk fields. As usual, we may choose to locate these sources at the boundary where they may be translated into sources for the local single-trace \mathcal{O}^i above. Now, single and multi-trace sources mix under time-evolution, but this mixing is again controlled by the $1/N$ expansion: since the stress tensor generates time evolution, to any order in $1/N$ a light operator $\mathcal{O}(x)$ can be replaced inside such correlators with an operator at another time that is a sum over k of k -trace terms weighted by $1/N^{k-1}$. Semi-classical behavior is preserved in time, so it should suffice to restrict attention to states of the form

$$|\alpha\rangle := U|0\rangle, \quad U := T e^{-i\alpha J}, \quad (2.3a)$$

$$J := \sum_{k=1} N^{-(k-1)} \int dx_1 \dots dx_k j_{i_1 \dots i_k}(x_1, \dots, x_k) \mathcal{O}^{i_1}(x_1) \dots \mathcal{O}^{i_k}(x_k), \quad (2.3b)$$

where the classical sources $j_{i_1 \dots i_k}$ are fixed smooth c -number functions of order N^0 and we allow distributional terms so that terms in the k -trace contributions to (2.3b) may effectively include only $m < k$ integrals over operator location.

We have introduced the real number α to be used as an expansion parameter. The symbol T denotes time-ordering, and we employ the convenient abuse of notation that defines $T e^{-i\alpha J}$ to be the natural time-ordered exponential associated with the particular representation of J given above as an integral over the CFT spacetime. Using the standard AdS/CFT dictionary, the above normalizations would give the bulk field ϕ^i dual to \mathcal{O}^i an expectation value $\langle \phi \rangle_\alpha \sim \alpha$.

We wish to choose α so that (2.3) would behave semi-classically in a bulk gravitational dual. Since the bulk is perturbative at large N , bulk quantum fluctuations become negligible in the limit $\alpha \gg 1$ and gravitational back reaction scales like

$$G\langle \phi \rangle_\alpha^2 \sim \frac{\alpha^2}{N^2}. \quad (2.4)$$

To allow this effect to be as large as possible consistent with a perturbative treatment we take

$$\alpha = \epsilon N \gg 1 \quad (2.5)$$

where ϵ is a small parameter to be held constant in the limit $N \rightarrow \infty$.

Our particular interest is in the effect of such sources on a region A of some Cauchy surface Σ in the CFT spacetime, or more generally on the associated domain of dependence $D(A)$. The region A is held fixed as we take $N \rightarrow \infty$. We denote the complementary region on this Cauchy surface by A^c , so that $\Sigma = A \cup A^c$. As a density matrix, the state $|\alpha\rangle$ is $\sigma_\alpha := |\alpha\rangle\langle\alpha|$. The associated reduced density matrices and the un-perturbed modular Hamiltonians for A and A^c are then

$$\begin{aligned}\rho_{A\alpha} &= \rho_A := \text{Tr}_{A^c}(\sigma_\alpha), & H_A &:= -\log(\rho_{A(\alpha=0)}), \\ \rho_{A^c\alpha} &= \rho_{A^c} := \text{Tr}_A(\sigma_\alpha), & H_{A^c} &:= -\log(\rho_{A^c(\alpha=0)}).\end{aligned}\quad (2.6)$$

As implied above, we will often suppress the label α on ρ_A, ρ_{A^c} when the meaning is clear, though H_A, H_{A^c} will always represent the modular Hamiltonians at $\alpha = 0$. An important property of these objects is

$$[H_A, H_{A^c}] = 0, \quad (2.7)$$

where both H_A and H_{A^c} are interpreted as operators on the full CFT Hilbert space by tensoring each with the identity operator ($\mathbb{1}_{A^c}$ or $\mathbb{1}_A$) on the complementary region. The result (2.7) would be immediate for truly bipartite systems that decompose as a tensor product of a system on A with a system on A^c . It is less trivial in quantum field theory, but still holds for ball-shaped regions A in CFTs since the vacuum is invariant under a Z_2 conformal symmetry that exchanges A and A^c . This symmetry must then exchange H_A and H_{A^c} as well, requiring their commutator to be symmetric. The manifest antisymmetry of $[H_A, H_{A^c}]$ then forces the commutator to vanish.

It is also useful to define the operator $K := H_A - H_{A^c}$. Recall that K annihilates the vacuum:

$$K|0\rangle = (H_A - H_{A^c})|0\rangle = 0. \quad (2.8)$$

For ball-shaped regions A in a constant-time slice, K generates the conformal isometry of $\mathbb{R} \otimes S^{d-1}$ that moves A orthogonally to itself while preserving the domain of dependence $D(A)$; i.e., it generates the natural forward “time-translation” on $D(A)$ and the natural backward “time-translation” on $D(A^c)$ [16]. For lack of a better name, we will refer to K as the boost Hamiltonian, though this term is only an accurate description of K in the large-sphere limit when A becomes a half-plane.

Finally, we introduce a basis for the space of light local gauge-invariant single trace operators $\{\mathcal{O}_A^i(x)\}$ ($\{\mathcal{O}_{A^c}^i(x)\}$) on the domain of dependence $D(A)$ ($D(A^c)$) that again satisfy (2.1) and (2.2). Below we will consider ball-shaped regions in a constant-time Cauchy surface $\Sigma = S^{d-1}$ so that the modular Hamiltonians H_A, H_{A^c} can be expressed as integrals of the stress tensor over A, A^c in any CFT [16].

2.1 Adapting the source

A key step in our argument will be to write the source-operator J from (2.3b) in a manner adapted to the decomposition $\Sigma = A \cup A^c$. The basic idea is to use the Heisenberg equations of motion for the CFT to express J in terms of operators on Σ . Thinking of the CFT Hilbert space as a tensor product of separate Hilbert spaces for A and A^c would then allow J to

be written as a sum of terms, each of which is the tensor product of a (possibly trivial) operator on A with one on A^c .

$$J = \sum_a J_A^a \otimes J_{A^c}^a. \quad (2.9)$$

We would like to then expand the $J_A^a, J_{A^c}^a$ as a sum or products of the $\{\mathcal{O}_A^i(x)\}$ ($\{\mathcal{O}_{A^c}^i(x)\}$) and use (2.1) to count powers of N .

There are, however, several potential obstacles to consider. First, evolving the source to the Cauchy surface Σ will generally mix the \mathcal{O}_i with both heavy operators and highly non-local expressions such as Wilson loops that cannot be expanded in terms of the $\{\mathcal{O}_A^i(x)\}$ ($\{\mathcal{O}_{A^c}^i(x)\}$). However, our calculations below will involve only correlation functions of sources with H_A, H_{A^c} . For the ball-shaped regions to be considered, up to the choice of zero-point our H_A, H_{A^c} are integrals of the stress tensor, which is one of our light operators. Since the stress tensor also generates time evolution, to any order in $1/N$ a light operator $\mathcal{O}(x)$ can be replaced inside such correlators with an appropriately-smeared sum of products of $\{\mathcal{O}_A^i(x)\}$ ($\{\mathcal{O}_{A^c}^i(x)\}$). So in this sense we may write

$$J = \mathcal{O}_A + \mathcal{O}_{A^c} + \sum_{k=2} \frac{1}{N^{k-1}} \int_{[D(A) \cup D(A^c)]^k} dx_1 \dots dx_k \tilde{j}_{i_1 \dots i_k}(x_1, \dots, x_k) \mathcal{O}^{i_1}(x_1) \dots \mathcal{O}^{i_k}(x_k), \quad (2.10)$$

where we have used the fact that operator mixing is suppressed by powers of $1/N$ and defined

$$\mathcal{O}_A = \int_{D(A)} j_i^A(x) \mathcal{O}_A^i(x), \quad \mathcal{O}_{A^c} = \int_{D(A^c)} j_i^{A^c}(x) \mathcal{O}_{A^c}^i(x). \quad (2.11)$$

The combinations of operators represented by $\mathcal{O}_A, \mathcal{O}_{A^c}$ will play important roles below. Although not explicitly indicated, we will make use of a similar decomposition of the multi-trace parts of (2.10) into sums of products of operators associated separately with A and A^c . Since (2.3) requires time-ordered products of such operators, we should really perform an expansion of the form (2.10) to express the source $J(t)$ at each time in terms of operators that evolve to the regions A, A^c , but to avoid clutter in our notation we will not explicitly indicate the time at which given operators are to act.

In terms of a holographic bulk dual, the rewriting of (2.3) as (2.10) could be described as evolving sources to a bulk Cauchy surface intersecting the boundary at $A \cup A^c$ and then using e.g. the methods of [28–34] to write the corresponding bulk operators in terms of boundary operators¹ in $D(A), D(A^c)$. At least for operators \mathcal{O}^i not associated with bulk gauge fields, this interpretation nicely side-steps issues (see e.g. [35–44]) associated with the expectation that our CFT will be a gauge theory so that its Gauss-law constraint forbids factorization of its Hilbert space into separate Hilbert spaces for A and A^c .

However, even without a Gauss-law constraint, quantum field theory Hilbert spaces do not admit a precise tensor product structure. This fact is associated with singularities in various n -point correlation functions which can give contributions localized precisely on

¹Since our analysis is perturbative, this construction may be performed in empty AdS so that ball-shaped regions A, A^c define Rindler-AdS regions with a common bifurcate Rindler horizon.

the boundary ∂A where A and A^c meet.² The issue could be ignored if we were to consider only correlators smeared with smooth functions; we could simply deform the smearing functions so that they vanish in some small neighborhood of ∂A and then recover the original undeformed correlators by continuity as these neighborhoods shrink to zero size. But the fact that H_A, H_{A^c} are integrals of the stress tensor against non-smooth functions³ means that more care will be required. In section 3 below we simply restrict to sources that induce changes in energy $\Delta\langle H_A \rangle$ and entropy ΔS_A that can be approximated by replacing each term in (2.10) with a source that differs from the original in some neighborhood A_{collar} of ∂A and vanishes smoothly in a smaller neighborhood $A_{\text{source-free}}$ of ∂A , and then letting the width of A_{collar} vanish. Roughly speaking, these are the sources that do not produce distributional terms localized at ∂A . We refer to such sources as having non-singular action on the modular Hamiltonian H_A .

In practice, rather than working through the above limit explicitly, we will simply consider the regulated operators mentioned above which we take to vanish in a common neighborhood $A_{\text{source-free}}$ of ∂A . The limit $A_{\text{collar}}, A_{\text{source-free}} \rightarrow \emptyset$ will be left implicit. Such sources may for example be obtained by choosing the original stress-tensor sources to be supported in the interior of $D(A) \cup D(A^c)$. Nevertheless, the fact that the multi-trace terms in either (2.3) or (2.10) are multi-local, means that they can include products of operators in $D(A)$ with those in $D(A^c)$. Thus U is generally not a product of separate unitary transformations on $D(A)$ and $D(A^c)$.

3 Results with no singular terms

We now study the changes in energy and entropy associated with applying U to $|0\rangle$ for the above sources.

3.1 Energy

Changes in the energy are straightforward to evaluate and take the form

$$\begin{aligned}\Delta\langle H_A \rangle &= \langle \bar{T} e^{i\alpha J} H_A T e^{-i\alpha J} - H_A \rangle \\ &= i\epsilon N \langle [J, H_A] \rangle - \frac{\epsilon^2 N^2}{2} \langle [J, [J, H_A]]_T \rangle + O(\epsilon^3 N^2),\end{aligned}\tag{3.1}$$

where \bar{T} denotes anti-time-ordering and $[J, [J, H_A]]_T$ is the operation defined by the second order expansion of the first line; as indicated by the notation, one may think of this term as an appropriately time-ordered version of a double commutator.

Since the first order term also clearly follows from the first line, it remains only to show that the omitted terms are of order $\epsilon^3 N^2$ and in particular involve no higher powers of N . We expect that this argument is also standard, but we state it here for completeness. The point is to note the repeated commutator structure of $\bar{T} e^{i\alpha J} H_A T e^{-i\alpha J} - H_A$, which requires that at order α^n all n sources must be connected to each other and also to H_A .

²In the bulk description mentioned above, it is associated with the fact that operators at the bifurcation surface of the Rindler horizon cannot be mapped to either the $D(A)$ or $D(A^c)$.

³The first derivative of the smearing function is discontinuous at ∂A .

Considering for the moment only the single-trace contribution $\mathcal{O}_A + \mathcal{O}_{A^c}$ to all sources J from (2.10) and using (2.1) then gives only terms of order $\epsilon^n N^2$.

One may then show that multi-trace contributions are further suppressed by at least an additional N^{-2} : since the coefficient of each multi-trace contribution to (2.10) contains an explicit factor of $1/N$, the only possible exception could come from including a single double-trace contribution $\mathcal{O}^{i_1}\mathcal{O}^{i_2}$ to one of the sources. But since $\langle\mathcal{O}^{i_1}\rangle = \langle\mathcal{O}^{i_2}\rangle = 0$, the repeated commutator structure again means that all that all non-zero contributions are fully connected. From (2.1), the replacement of a single-trace operator by $\mathcal{O}^{i_1}\mathcal{O}^{i_2}$ thus yields an extra $1/N$ in this connected correlator giving the stated suppression by two powers of $1/N$.

We should also comment further on the linear term in (3.1). Recall that we take sources in (2.10) to vanish near ∂A . This must in particular be true of the single-trace term, which can thus be written as $\mathcal{O}_A + \mathcal{O}_{A^c}$. Writing H_A as an integral of $N\mathcal{O}^0$ (plus a constant) and using (2.1) would suggest that this single-trace part makes the linear term of order ϵN^2 . But the \mathcal{O}_{A^c} term clearly commutes with H_A . And since $[\mathcal{O}_A, H_{A^c}] = 0$ we have $[\mathcal{O}_A, H_A] = [\mathcal{O}_A, K]$. The fact that K annihilates the vacuum then implies $\langle[\mathcal{O}_A, H_A]\rangle = 0$. As a result, the linear term in (3.1) receives contributions only from multi-trace terms in (2.10) and is thus suppressed by an extra N^{-2} as described above. Thus the linear term is in fact of order ϵN^0 .

3.2 Entropy

Computing changes in entropy and relative entropy is more complicated than computing $\Delta\langle H_A\rangle$. The main issue is that expanding the logarithm in S_A requires using the highly non-trivial Baker-Campbell-Hausdorff (BCH) formula. Nevertheless, subject to the assumption that all sources in (2.10) vanish near ∂A , we will argue below that changes in the entropy S_A and relative entropy R_A take the form

$$R_A := \Delta\langle H_A\rangle - \Delta S_A = \langle e^{H_A}\Delta\rho_A[f(K) - f'(K)]\Delta\rho_A e^{H_A}\rangle + O(\epsilon^3 N^2), \quad (3.2a)$$

$$= \epsilon^2 N^2 \langle \mathcal{O}_A K \mathcal{O}_A \rangle + O(\epsilon^3 N^2) + O(\epsilon^2 N^0) \quad (3.2b)$$

$$\Delta S_A = O(\epsilon N^0) \quad (3.2c)$$

where $\Delta\rho_A = \rho_A - \rho_{A0}$ and f is the smooth function

$$f(y) = \frac{y}{1 - e^{-y}}. \quad (3.3)$$

This f appears because it is the generating function of the Bernoulli numbers which play an important role in the BCH formula [45]. While it is generally redundant to describe $\langle H_A\rangle$, S_A , and $R_A := \langle H_A\rangle - S_A$, we choose to do so — and in fact give two expressions for R_A — both in order to state definite results and to describe the relative sizes of various contributions. For example, the explicit term in (3.2a) is a general result for the second order term in the relative entropy R_A of any bipartite quantum system,⁴ but the large N

⁴While continuum field theories are not truly bipartite, we may nevertheless use this formula due to our assumption that sources vanish near ∂A . See appendix A for details.

structure is more apparent from (3.2b). Since our current interest focuses on the latter, and since (3.2b) may be derived from more general arguments, we relegate the calculation leading to (3.2a) to appendix A. Equation (3.2a) may nevertheless be useful for applications that require subleading terms of order $\epsilon^2 N^0$ as well as the leading term of order $\epsilon^2 N^2$.

Although deriving these results will require some work, the intuition behind (3.2b) and (3.2c) is easy to understand. Note that keeping only the single-trace term would make U a product of separate (commuting) unitary transformations U^A and U^{A^c} on A and A^c . This would require $\Delta S_A = 0$ and thus $R_A = \Delta \langle H_A \rangle$. So contributions to ΔS_A require the multi-trace terms in (2.10), which are of order N^0 . Deriving (3.2c) thus amounts to controlling cross-terms involving both single-trace and multi-trace source terms. This is done in section 3.3 below.

The explicit term in (3.2b) is the second-order effect of purely single-trace sources on R_A , or equivalently on $\langle H_A \rangle$. Since either can depend only on the restriction of the state to A , they are unchanged by the action of the U^{A^c} defined above. We may thus consider only U^A , which commutes with H_{A^c} and thus induces identical changes in both H_A and K . Indeed, since the linear term in (3.1) vanishes for single-trace sources that vanish near ∂A , it suffices to compute only the second-order term

$$-\frac{\epsilon^2 N^2}{2} \langle [\mathcal{O}_A, [\mathcal{O}_A, K]]_T \rangle. \quad (3.4)$$

Using $K|0\rangle = \langle 0|K = 0$ then shows the only non-zero term to be the one displayed in (3.2b). Along with the fact that the first order change in R_A vanishes identically by the first law, the errors in (3.1) and (3.2c) then imply those in (3.2b).

3.3 Multi-trace contributions to ΔS_A

The task that remains is to show that including multi-trace contributions to (2.10) can change S_A only by terms of order N^0 . The argument is somewhat lengthy, so we break it into several parts. We first reorganize the action of the single-trace source-terms in order to show that they have little impact. We then work to write a series expansion of S_A to which we can usefully apply the large- N counting rule (2.1). Here there are two difficulties, one of which is associated with the fact that S_A is defined as a non-linear function of the reduced density matrix ρ_A on region A and not as a vacuum expectation value. Any power series expansion of S_A thus naturally involves many traces over A_c that must be eliminated in order to use (2.1). The other involves controlling contributions from possible disconnected correlators. After completing these tasks, we combine the results and count powers of N .

3.3.1 Reorganizing the action of single-trace sources

To begin, recall the definitions

$$\rho_A = \text{Tr}_{A^c} \sigma_\alpha = \text{Tr}_{A^c} (T e^{-i\alpha J} \sigma_0 \bar{T} e^{i\alpha J}). \quad (3.5)$$

Recall also that dropping the multi-trace contributions and computing only

$$U_{\text{single}} = T e^{-i\alpha \int dt J_{\text{single}}} \quad (3.6)$$

would give a product of separate unitary transformations on A and A^c that do not change S_A . Similarly, we may note that the entropy of ρ_A must be identical to the entropy of

$$\rho_A^{\text{conj}} = \text{Tr}_{A^c} (U_{\text{single}}^{-1} \sigma_\alpha U_{\text{single}}) = [U_{\text{single}}^A]^{-1} \rho_A U_{\text{single}}^A, \quad (3.7)$$

where U_{single}^A is the part of U_{single} that acts on A . We thus have

$$S_A = -\text{Tr}_A (\rho_A^{\text{conj}} \log \rho_A^{\text{conj}}). \quad (3.8)$$

Now, σ_α itself is defined (see (2.3)) by conjugating $|0\rangle\langle 0|$ with a time-ordered unitary U . This unitary is the product of many factors of the form

$$e^{-i\alpha J(t)\delta t} = e^{-i\alpha J_{\text{single}}(t)\delta t} e^{-i\alpha J_{\text{multi}}(t)\delta t} + O(\delta t^2) \quad (3.9)$$

where we have separated the single- and multi-trace parts of the source at time t . Note that our new conjugation by U_{single} merely replaces the U in (2.3) by $U_{\text{single}}^{-1}U$, or equivalently replaces each factor (3.9) with

$$e^{-i\alpha J_{\text{multi}}^{\text{conj}}(t)\delta t}, \quad (3.10)$$

where $J_{\text{multi}}^{\text{conj}}(t)$ is a multi-trace term conjugated by a time-ordered exponential built from single-trace sources at earlier times; there is no effect on (3.9) from sources at later times as these merely cancel between U_{single}^{-1} and U . One may make an analogy between U and the time-ordered exponential that implements Heisenberg-picture time-evolution in some quantum system, with J_{single} playing the role of the free Hamiltonian and J_{multi} playing the role of the interaction terms. Our conjugation by U_{single} then plays the role of passing to the interaction picture, where the new time evolution is a product of factors like (3.10).

We wish to write

$$\rho_A^{\text{conj}} = \rho_{A0} (1 + \rho_{A0}^{-1} \Delta \rho_A^{\text{conj}}) \quad (3.11)$$

for

$$\Delta \rho_A^{\text{conj}} = \rho_A^{\text{conj}} - \rho_{A0} = \text{Tr}_{A^c} (U_{\text{single}}^{-1} \sigma_\alpha U_{\text{single}} - \sigma_0), \quad (3.12)$$

and to expand (3.8) in powers of $\rho_{A0}^{-1} \Delta \rho_A^{\text{conj}}$. We then expand $\Delta \rho_A^{\text{conj}}$ in powers of α . In this latter step will write each $\Delta \rho_A^{\text{conj}}$ as a sum of terms of the form

$$\text{Tr}_{A^c} ([J_{\text{multi-1}}^{\text{conj}}, [J_{\text{multi-2}}^{\text{conj}} \cdots [J_{\text{multi-k}}^{\text{conj}}, |0\rangle\langle 0|] \cdots]), \quad (3.13)$$

where $J_{\text{multi-j}}^{\text{conj}}$ is a time integral of $J_{\text{multi}}^{\text{conj}}(t)$ and the factors are appropriately time-ordered.

It is important to note that $|0\rangle\langle 0|$ is not conjugated by any such unitary; the privileged position of this operator at the end of the chain of repeated commutators means that, in our analogy with transforming between the Heisenberg and interaction pictures, this operator can be thought of as labeled by the earliest possible time so that the picture-changing transformation acts on it trivially. The point of the form (3.13) is that we will shortly (see section 3.3.2) transform the expansion of (3.8) into a form involving only vacuum correlators of products of operators. The repeated commutator structure of $J_{\text{multi-j}}^{\text{conj}}$ will then forbid single-trace terms from appearing in correlators unless they are appropriately connected to multi-trace terms.

3.3.2 The entropy as a correlator

The next step in our argument is to show how S_A can be written as a sum of vacuum correlators of products of the \mathcal{O}^i with functions of the boost Hamiltonian K , and where the remaining N -dependence of each term follows directly from powers of $\alpha = \epsilon N$ and (2.10). In particular, all extra traces over A^c will be removed, and the modular Hamiltonians H_A , H_{A^c} will not appear except in the combination $K = H_A - H_{A^c}$. This last feature will be critical in controlling contributions from disconnected correlators.

As a first step toward this goal, we may use (3.11) to express S_A as the vacuum correlator

$$\begin{aligned} S_A &= -\text{Tr}_A [\rho_{A0} (1 + \rho_{A0}^{-1} \Delta \rho_A^{\text{conj}}) \log \rho_A^{\text{conj}}] \\ &= -\langle (1 + \rho_{A0}^{-1} \Delta \rho_A^{\text{conj}}) \log \rho_A^{\text{conj}} \rangle, \end{aligned} \quad (3.14)$$

where operators on region A are to be interpreted as operators on the full CFT Hilbert space by tensoring them with the identity on A^c . Note that $\log(\rho_A^{\text{conj}} \otimes \mathbb{1}_{A^c}) = (\log \rho_A^{\text{conj}}) + \log \mathbb{1}_{A^c} = (\rho_A^{\text{conj}})$, so that we may interpret $\log \rho_A^{\text{conj}}$ as the logarithm of an operator on the full Hilbert space.

We next remove the explicit traces over A^c . These enter through the definition (2.6), and are potentially problematic because the sources involve products of operators on A with operators on A^c . The trick to proceeding is to use the assumption that each term in (2.10) is supported away from ∂A (so that contributions to each source-term from A and A^c commute with each other) to write each term (3.13) as a sum of terms in which all operators on A^c have been commuted to act directly either on $|0\rangle$ from the left or on $\langle 0|$ from the right.

Using the entanglement properties of $|0\rangle$, we may now replace each operator in A^c by a so-called “mirror operator” on A ; see e.g. [46], though the particular terminology is from the more recent [47]. To explain how this works, let us for the moment take A to be the “southern” hemisphere of our S^{d-1} at $t = 0$. In any relativistic theory, given an operator \mathcal{O} , not necessarily a scalar or even local, we may study the CPT conjugate operator \mathcal{O}^{CPT} . Since the parity operation exchanges the north and south hemispheres, the CPT conjugate of any northern hemisphere operator at $t = 0$ (i.e., on A^c) is an operator on the $t = 0$ southern hemisphere A . Furthermore, these operators satisfy

$$\mathcal{O}^{\text{CPT}}|0\rangle = e^{K/2}\mathcal{O}|0\rangle; \quad \langle 0|\mathcal{O}^{\text{CPT}\dagger} = \langle 0|\mathcal{O}^\dagger e^{K/2}, \quad (3.15)$$

where the second relation is just the Hermitian conjugate of the first. These are just Kubo-Martin-Schwinger (KMS) relations in terms of the imaginary time evolution generated by our K , so they encode the thermal nature of $|0\rangle$ with respect to H_A . Thus the mirror operator $\tilde{\mathcal{O}} := e^{-K/2}\mathcal{O}^{\text{CPT}}e^{K/2}$ satisfies

$$\tilde{\mathcal{O}}|0\rangle = \mathcal{O}|0\rangle; \quad \langle 0|\tilde{\mathcal{O}}^\dagger = \langle 0|\mathcal{O}^\dagger e^K. \quad (3.16)$$

The key point is that the CPT conjugate of an \mathcal{O}^i is just another \mathcal{O}^j (or perhaps a linear combination thereof). So at the expense of introducing additional factors of

$e^{-K/2} = e^{-H_A/2} e^{H_{A^c}/2}$, we may replace the A^c operators acting on σ_0 by CPT-conjugate operators acting on A . Conformal invariance then guarantees that we can again perform a corresponding operation to replace operators on A^c with those on A for more general ball-shaped regions. The new operators $e^{H_{A^c}/2}$ may then be commuted past A -operators to act on σ_0 , where they may be replaced by factors of $e^{H_A/2}$ using (2.8). The net result is thus to write $\Delta\rho_A$ as a sum of terms of the form

$$\text{Tr}_{A^c} (\mathcal{O}_A^{i_1} \dots \mathcal{O}_A^{i_n} e^{-H_A/2} \mathcal{O}_A^{j_1} \dots \mathcal{O}_A^{j_m} e^{+H_A/2} |0\rangle\langle 0| e^{+H_A/2} \mathcal{O}_A^{k_1} \dots \mathcal{O}_A^{k_r} e^{-H_A/2} \mathcal{O}_A^{l_1} \dots \mathcal{O}_A^{l_s}). \quad (3.17)$$

Operators supported away from A^c can now be pulled outside the trace over A^c . By assumption these include all \mathcal{O}_A^i , but we should take care with the factors of $e^{H_A/2}$ which include support near ∂A . We may do so in each term by using the Zassenhaus formula

$$e^{X+Y} = e^X e^Y e^{-[X,Y]/2} e^{\frac{1}{6}(2[Y,[X,Y]]+[X,[X,Y]])} \dots \quad (3.18)$$

for $H_A/2 = X + Y$, with $X = H_A^{\text{far}}/2$ a Hermitian integral of the subtracted stress tensor $N\mathcal{O}_A^0$ weighted by a smooth function supported away from ∂A and Y a similar integral supported close enough to ∂A to avoid overlap with the support of any \mathcal{O}_A^i appearing in the given term. Since both are smooth integrals, the supports of X and Y will overlap with each other and thus yield non-zero commutators in (3.18). It is convenient to collect all factors involving Y into a single operator $e^{H_A^{\text{near}}/2}$. Note that commutators of X, Y receive contributions only from regions in the support of both operators, so that $e^{H_A^{\text{near}}/2}$ is again supported in a region close to ∂A that avoids overlap with the support of any \mathcal{O}_A^i appearing in the given term.

We now have

$$e^{-H_A/2} = e^{-H_A^{\text{far}}/2} e^{-H_A^{\text{near}}/2}, \quad (3.19a)$$

$$e^{+H_A/2} = e^{+H_A^{\text{near}}/2} e^{+H_A^{\text{far}}/2} \quad (3.19b)$$

$$e^{-H_A/2} = e^{-H_A^{\text{near}\dagger}/2} e^{-H_A^{\text{far}}/2}, \quad (3.19c)$$

$$e^{+H_A/2} = e^{+H_A^{\text{far}}/2} e^{+H_A^{\text{near}\dagger}/2}, \quad (3.19d)$$

where (3.19b) is the inverse of (3.19a) and the final two relations (3.19c), (3.19d) are the adjoints of (3.19a), (3.19b). The expressions (3.19) allow us to safely reformulate (3.17) as

$$\begin{aligned} & \mathcal{O}_A^{i_1} \dots \mathcal{O}_A^{i_n} e^{-H_A^{\text{far}}/2} \mathcal{O}_A^{j_1} \dots \mathcal{O}_A^{j_m} \\ & \times \text{Tr}_{A^c} (e^{-H_A^{\text{near}}/2} e^{+H_A^{\text{near}}/2} e^{+H_A^{\text{far}}/2} |0\rangle\langle 0| e^{+H_A^{\text{far}}/2} e^{+H_A^{\text{near}}/2} e^{-H_A^{\text{near}}/2}) \\ & \times \mathcal{O}_A^{k_1} \dots \mathcal{O}_A^{k_r} e^{-H_A^{\text{far}}/2} \mathcal{O}_A^{l_1} \dots \mathcal{O}_A^{l_s}. \end{aligned} \quad (3.20)$$

Since $e^{-H_A^{\text{near}}/2}$ is the inverse of $e^{H_A^{\text{near}}/2}$, these operators all cancel. The remaining factors of $e^{+H_A^{\text{far}}/2}$ may then be extracted from the trace as well. Reversing the steps involving H_A^{near} , H_A^{far} then reconstructs the original factors of $e^{\pm H_A/2}$ outside the trace, leaving just (3.17) where the trace acts only on $|0\rangle\langle 0|$ as one would naively expect. Using $\text{Tr}_{A^c} |0\rangle\langle 0| = e^{-H_A}$ then removes all explicit traces over A^c . While we could take this final factor of e^{-H_A} to

cancel the two factors of $e^{H_A/2}$ adjacent to $|0\rangle\langle 0|$ in (3.17), due to the structure of (3.13) it is convenient to instead keep all of these factors explicit. The latest factor of e^{-H_A} coming from the trace is distinguished, and we shall refer to it below as the ‘final’ e^{-H_A} .

The above steps have written each factor of $\Delta\rho_A^{\text{conj}}$ as a product of \mathcal{O}_A^i with two factors of $e^{-H_A/2}$, two factors of $e^{+H_A/2}$, and our final e^{-H_A} . But as noted above we wish to expand in powers of $\rho_{A0}^{-1}\Delta\rho_A^{\text{conj}} = e^{H_A}\Delta\rho_A^{\text{conj}}$. We may then use

$$\mathbb{1} = e^{H_{Ac}} e^{-H_{Ac}} e^{-H_{Ac}/2} e^{-H_{Ac}/2} e^{H_{Ac}/2} e^{H_{Ac}/2} \quad (3.21)$$

to insert the indicated factors of $e^{H_{Ac}/2}$. Commuting them past the \mathcal{O}_A^i and exponentials of H_A as needed allows us to replace all exponentials of H_A with exponentials of $K = H_A - H_{Ac}$. In particular, in this way we obtain a distinguished ‘final’ factor of e^{-K} .

We are now close to achieving our goal. If $\rho_{A0}^{-1}\Delta\rho_A^{\text{conj}}$ would commute with ρ_{A0} , we could use $\log \rho_A^{\text{conj}} - \log \rho_{A0} = \log(1 + \rho_{A0}^{-1}\Delta\rho_A^{\text{conj}})$ and expand (3.14) in a standard Taylor series. Combined with our results above, this would give an infinite sum of terms with each being a vacuum correlator of a product of \mathcal{O}_i and exponentials of K , consistent with the form required above.⁵

But since $\rho_{A0}^{-1}\Delta\rho_A^{\text{conj}}$ and ρ_{A0} generally do not commute we must use the Baker-Campbell-Hausdorff formula to evaluate $\log \rho_A^{\text{conj}} = \log[\rho_{A0}(1 + \rho_{A0}^{-1}\Delta\rho_A^{\text{conj}})]$. An explicit formulation of this identity due to Dynkin takes the form (see e.g. [48])

$$\log(e^X e^Y) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{p_i+q_i \geq 1} \frac{[X^{(p_1)}Y^{(q_1)} \dots X^{(p_k)}Y^{(q_k)}]}{(\sum_{i=1}^k (p_i + q_i)) (\prod_{i=1}^k p_i! q_i!)} \quad (3.22)$$

where

$$[X^{(p_1)}Y^{(q_1)} \dots X^{(p_k)}Y^{(q_k)}] \equiv \underbrace{[X, [X, \dots [X, \underbrace{[Y, \dots [Y, \underbrace{[X, \dots [X, \underbrace{[Y, [Y, \dots Y] \dots]}_{q_k \text{ } Y\text{s}}]}_{p_k \text{ } X\text{s}}]}_{q_1 \text{ } Y\text{s}}]}_{p_1 \text{ } X\text{s}}] \dots]}_{q_k \text{ } Y\text{s}}] \dots] \quad (3.23)$$

We wish to set $X = -H_A = \log \rho_{A0}$ and $Y = \log(1 + \rho_{A0}^{-1}\Delta\rho_A^{\text{conj}})$.

The important observation is that the form of $\rho_{A0}^{-1}\Delta\rho_A^{\text{conj}}$ found above clearly commutes with H_{Ac} , so $[H_{Ac}, Y] = 0$ as well. Recalling from (2.7) that $X = -H_A$ also commutes with H_{Ac} allows us to replace each X in the above repeated commutators with $-K = H_{Ac} - H_A$. The result is an expression for S_A as an infinite sum over terms, each of which is a vacuum correlator involving only products of \mathcal{O}_A^i and functions of K , and with all further explicit dependence on N coming from (2.10) as desired.

3.3.3 Counting powers of N

We are now ready to count powers of N . Consider then contributions at order α^p involving r single-trace operators $\mathcal{O}_{\text{from single}}$ that come from single-trace source-terms and s single-trace operators $\mathcal{O}_{\text{from multi}}$ that come from multi-trace source-terms. Such a contribution is equal to a product of connected correlators each with r_i of the $\mathcal{O}_{\text{from single}}$ and s_i of the

⁵With the exception of the term $\langle \rho_{A0}^{-1}\Delta\rho_A^{\text{conj}} H_A \rangle$. This term resembles the multi-trace contributions to $\Delta\langle \log \rho_{A0} \rangle$ considered in section 3.1 and is also of order N^0 for similar reasons.

$\mathcal{O}_{\text{from multi}}$ such that $r = \sum_i r_i$ and $s = \sum_i s_i$. We need not keep track of the factors of K since each contributes an explicit factor of N that is cancelled by the extra $1/N$ from (2.1) associated with the \mathcal{O}^0 it contributes to any correlator. Each connected correlator gives $r_i + s_i - 2$ factors of $1/N$ by (2.1). The single-trace sources provide an additional r explicit powers of N . For double-trace sources the explicit N from $\alpha = \epsilon N$ cancels against the $1/N$ in (2.10) and for higher-trace sources the contribution is more suppressed. So the total number of factors of $1/N$ is greater than or equal to $\sum_i (s_i - 2)$. We wish to show that this sum is non-negative.

Recall that the reason multi-trace sources affect the entropy is that they can contain products of operators in A with operators in A^c . But a given term will only be sensitive to the correlations created by these sources if at least one operator from A and one from A^c appear in the same connected correlator. So it is natural to expect that the entropy will only receive contributions from terms with $s_i \geq 2$ for all i , which would imply the desired result.

To show that this is so, we note that there are no contributions from terms with $s_i = 0$. This is because all of the $\mathcal{O}_{\text{from single}}$ occur in nested commutators inside some $J_{\text{multi-}j}^{\text{conj}}$ and therefore must be connected to at least one $\mathcal{O}_{\text{from multi}}$ in order to contribute.

Now suppose $s_i = 1$, where we take the relevant $\mathcal{O}_{\text{from multi-}j}$ to come from a m -trace source term living inside some $J_{\text{multi-}j}^{\text{conj}}$. Since $\langle \mathcal{O}_{\text{from multi-}j} \rangle = 0$, a non-vanishing connected correlator must involve other operators. None of these can come from single-trace source terms in other $J_{\text{multi-}k}^{\text{conj}}$ for $k \neq j$, which must instead stay attached to their own multi-trace sources. So since $s_i = 1$, if any of these operators come from a final K then the correlator must vanish due to $K|0\rangle = 0 = \langle 0|K$ and the fact that (3.13) prohibits a final K from intervening between factors coming from any given $J_{\text{multi-}j}^{\text{conj}}$. So the correlator consists only of $\mathcal{O}_{\text{from multi-}j}$, single-trace source-terms, and K 's coming from a single $J_{\text{multi-}j}^{\text{conj}}$. Terms of this form exponentiate, and have the effect of replacing the operator $\mathcal{O}_{\text{from multi-}j}$ in $J_{\text{multi-}j}^{\text{conj}}$ with a classical source of order N , demoting $J_{\text{multi-}j}^{\text{conj}}$ to an $(m-1)$ -trace source of order N^{-m-3} (i.e., the same order as the $(m-1)$ -trace sources already appearing in (2.10)). So we need not consider such terms separately when counting powers of N . In particular, double-trace sources demoted to single-trace sources in this way can be absorbed into the U_{single} of section 3.3.1 and do not contribute to the entropy. So all relevant terms have $s_i \geq 2$ and the largest possible contribution to the entropy is of order N^0 as anticipated above.

4 Discussion

The work above contains first steps toward studying the von Neumann entropy of excited states in CFTs satisfying the 't Hooft large N counting rule (2.1). We considered the entropy of ball-shaped regions for states produced by real-time sources with non-singular action on the modular Hamiltonian and of the form (2.10). From a dual bulk point of view, such sources produce small classical waves — or more properly quantum coherent states with large amplitude of order ϵN — on both sides of a Rindler horizon but which do not disturb the bifurcation surface itself. They also add $O(N^0)$ entangled sets of particles

across the horizon. As the former does not change the HRT entropy, one expects that the CFT entanglement changes only at order N^0 . We have verified that this is indeed the case by a direct argument in the CFT. While our results are directly formulated in terms of a $1/N$ expansion, it may be interesting to follow [5] and attempt to formulate a version of our results that would hold in arbitrary CFTs, whether or not they have a large- N counting rule like (2.1).⁶

One consequence of our work is the explicit formula (3.2b) for the order $\epsilon^2 N^2$ relative entropy. When interpreted in the bulk theory, this formula is precisely the bulk stress tensor on the A side of the bulk horizon, integrated so as to give the associated change in the boost energy K . This may be seen from the fact that (3.2b) is just the contribution to K from the excitations on the A side of the bifurcation surface, and by bulk causality that it thus gives the value of the boost Hamiltonian K in a related bulk solution produced from the vacuum using only the sources \mathcal{O}_A in the causal past of A . The above claim then follows by writing K in terms of a standard ‘bulk stress tensor’ that includes quadratic contributions from gravitons as well. For concreteness, one may choose the graviton contribution to be given by the AdS analogue of the Landau-Lifshitz ‘pseudo-tensor’ expression [49], though our requirement that perturbations vanish near the bifurcation surface means that many other choices give equivalent results.⁷

The above observation means that our analysis provides a new argument for the universal coupling of gravity to all classical forms of stress-energy. Though it differs in detail from [4], this interpretation of our classical ($O(N^2)$) result is inspired by the quantum ($O(N^0)$) argument of that reference and reinforces the connection found there between the universal structure of variations in S_A and the universal coupling of bulk gravity.

Another consequence is to give a derivation of the order N^0 Faulkner-Lewkowycz-Maldacena correction [23] to the Ryu-Takayagi and Hubeny-Rangamani-Takayanagi conjectures (for ball-shaped regions A). In the semi-classical bulk, this argument is perturbative in departures from empty AdS. It thus complements the original reasoning in [23] in that it does not rely on the Lewkowycz-Maldacena argument [12], or on any other use of the replica trick. At first order in ϵ the result follows from the first law, but the arguments of section 3 also give this result at higher orders. In general, we may decompose S_A into a first-law-piece and R_A . Recall that at second order R_A is quadratic in the first order contribution to $\Delta\rho_A$ and in particular is given by

$$-i\alpha \operatorname{Tr}_{A^c}([J_{\text{double}}, |0\rangle\langle 0|]) = \operatorname{Tr}_{A^c}(e^{-i\alpha J_{\text{double}}} |0\rangle\langle 0| e^{i\alpha J_{\text{double}}} - |0\rangle\langle 0|) + O(\epsilon^2), \quad (4.1)$$

where J_{double} is the double-trace part of (2.10). As before the only terms in J_{double} that can contribute to ΔS_A are those that involve one operator in A and one in A^c . Such terms throw entangled pairs into the bulk, with one member of each pair on each side of the bulk horizon. Given the agreement of bulk and CFT vacuum correlators, it is manifest that the

⁶We thank Tom Faulkner for discussion on this point.

⁷There is, however, a preferred expression when perturbations do not vanish at the bifurcation surface. The bulk relative entropy $H_A - S$ of the classical bulk is then the so-called canonical energy density of [50], which is naturally expressed as an integral of what one may call a bulk stress tensor. See [51] for further discussion of this connection.

corresponding change in bulk entanglement at this order can be computed just as we have done in the CFT and that the results agree. A similar argument holds at higher orders in ϵ , where the repeated commutator structure of (3.13) reproduces the effect of propagating small effects from the multi-trace sources through the large semi-classical coherent state produced by the single-trace parts of (2.10). It also extends to higher orders in N to argue that bulk entanglement gives the full series of $1/N$ corrections to HRT. Note, however, that since our sources are confined to $D(A) \cup D(A^c)$ all perturbations vanish at the bulk HRT surface. Our derivation is thus insensitive to possible perturbative shifts of the HRT surface of the kind predicted in [52]. Conversely, extending our results to allow sources supported on ∂A would in principle allow us to test the conjecture of [52] that the full CFT entropy is given by the generalized entropy of a quantum extremal surface, defined as the bulk surface that extremizes the bulk generalized entropy.

We have focussed on ball-shaped regions for simplicity, but we expect our arguments to generalize to arbitrary regions A . Indeed, as described in [53], one may address perturbative deformations of ball-shaped regions by inserting additional factors of the stress tensor (or, equivalently, of $N\mathcal{O}^0$). These insertions tend to add another operator to each connected correlator, giving an extra $1/N$ from (2.1) that cancels the explicit new factor of N . A similar argument can be used to compute corrections to (3.15) and construct the relevant “mirror operators” for these deformed regions. So the only ingredients of our analysis that remain to be checked are that H_A , H_{A^c} can be thought of as operators on the full CFT Hilbert space that commute both with each other, and with local operators supported away from ∂A . It should be possible to analyze these assumptions perturbatively as well. We expect that these will indeed hold at this level, but verifying this is beyond the scope of our work.

The extension to include sources supported near ∂A would clearly be of great interest. At linear order in ϵ the contribution to S_A from such sources is governed by the first law and was effectively studied in [2, 3]. But at this order there is no displacement of the bulk extremal surface. In particular, for ball-shaped regions A the bulk HRT surface continues to coincide with the bifurcation surface of the corresponding event horizon; i.e., with the causal holographic information surface of [54], which is expected to compute some coarse-grained version of the CFT entropy [54–56]. In contrast, the two are distinguished at second order, so comparing bulk and CFT computations may give insight into the nature of the relevant coarse-graining. In particular, one might hope to either support or falsify the conjecture [56] that it corresponds to the maximizing the entropy over all states for which certain one-point functions coincide with the original state.

In addition, comparing bulk and CFT computations to second order would derive or falsify the HRT conjecture at a non-trivial level. Even for static perturbations the fact that it avoids the replica trick would make this a useful complement to the Lewkowycz-Maldacena argument [12], and the perturbative method should be able to address general time-dependence to which [12] does not apply. And since second-order results are sensitive to the displacement of the HRT surface, they could in particular detect any possible motion of this surface in imaginary directions within the complexified AdS spacetime; i.e., they could help diagnose the possible role of complex extremal surfaces as explored in [14].

One reason that we have avoided such singular terms here is that they are in principle sensitive to the particular way that entanglement is to be defined in the CFT. Since the CFT is a gauge theory, this can involve a number of subtle issues [35–44]). Lewkowycz-Maldacena [12] suggests that the correct notion of CFT entropy is defined by the replica trick, which is precisely the computational tool we wish to avoid. However, the results of [35–44]) also suggest that the various definitions of entropy differ only by a local boundary term that will cancel in computing the mutual information between pairs of regions A, B . One should thus be able to ignore such concerns in this context. We hope to compute the ‘singular’ second order terms and to explore the above issues in the near future, perhaps using a suitably-generalized version of the calculation in appendix A.

Acknowledgments

It is a pleasure to thank David Berenstein, William Donnelly, Eric Dzienkowski, Monica Guica, Tom Hartman, Veronika Hubeny, Ted Jacobson, Nima Lashkari, Juan Maldacena, Mukund Rangamani, Vladimir Rosenhaus, Mark Van Raamsdonk, and Aron Wall for helpful discussions and feedback. We especially thank Tom Faulkner for his comments on an early draft of this paper. This work was supported by the National Science Foundation under grant numbers PHY12-05500 and PHY15-04541, and by funds from the University of California. In addition, K.K. is supported by the NSF GRFP under Grant No. DGE-1144085. D.M. thanks the Aspen Center and its NSF Grant #1066293 for their hospitality during the discussions where certain aspects of this project were conceived. He also thanks the KITP for their hospitality during the final stages of the project, where his work was further supported in part by National Science foundation grant number PHY11-25915.

A Computing the relative entropy

This appendix derives an explicit formula for the second order change $\delta^2 R_A$ in the relative entropy R_A corresponding to an arbitrary change $\delta\rho_A$ in the reduced density matrix for A . This result is not directly used in the main text, other than writing (3.2a). Our final expression (A.14) bears a striking resemblance to eq. (C.7) of [5]. In fact, (A.14) can also be derived from a straightforward generalization of the calculation leading to (C.7). We present a different, somewhat more involved calculation here because we hope that this approach will be useful for analyzing the additional terms at order ϵ^2 that arise when the sources in (2.10) do not vanish in a neighborhood of ∂A .

The setting for the calculation below is an arbitrary bipartite quantum system, meaning that it is the tensor product of a system on A and one on A^c . As discussed in the main text, the actual system we study is not strictly of this form, though it can be treated as such at least under our assumption that each term in (2.10) vanishes in a neighborhood of ∂A . So a critical step in realizing the hope expressed in the paragraph above is understanding modifications that arise when this assumption fails.

A.1 The Baker-Campbell-Hausdorff formula

We will compute $\delta^2 R_A$ using the BCH formula in the form (3.22). Since we work only to order Y^2 , it will be useful to rewrite (3.22) as

$$\log(e^X e^Y) = X + Y + \sum_{n=1}^{\infty} C_n [X^{(n)} Y^{(1)}] + \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} D_{n,k} [X^{(k)} Y^{(1)} X^{(n-k)} Y^{(1)}] + O(Y^3), \quad (\text{A.1})$$

where C_n and $D_{n,k}$ are rational numbers. It was shown in [45] that

$$C_n = \frac{(-1)^n B_n}{n!}, \quad (\text{A.2})$$

where B_n are the first Bernoulli numbers ($B_0 = 1, B_1 = -1/2, \dots$). The Bernoulli numbers can be defined by the smooth generating function

$$f(y) := \frac{y}{1 - e^{-y}} = \sum_{n=0}^{\infty} \frac{B_n}{n!} (-y)^n, \quad (\text{A.3})$$

where the sum on the right hand side converges for $y \in (-2\pi, 2\pi)$. Note that $B_{2m+1} = 0$ for $m \geq 1$.

To compute $\delta^2 R_A$ we will also need the coefficients $D_{n,0}$. To our knowledge these coefficients have not previously been explicitly computed. But a straightforward application of the recursive technique developed in [45] yields⁸

$$D_{n,0} = -\frac{1}{2(n+1)!} \sum_{k=1}^{n+1} (-1)^{n-k} \binom{n+1}{k} B_k B_{n+1-k} = -\frac{B_{n+1}}{2(n!)}. \quad (\text{A.5})$$

To obtain the second equality first note that for even n the only non-vanishing terms in the sum are $k = 1$ and $k = n$, and that these terms cancel. For $n = 1$ the equality is easily checked by hand. For odd $n \geq 3$, only the even k terms survive so $(-1)^{n-k} = -1$. The remaining sum is easily evaluated using a well known identity due to Euler and independently rediscovered by Ramanujan (see for example eq. (1.2) in [57])

$$\sum_{j=1}^m \binom{2m}{2j} B_{2j} B_{2m-2j} = -2m B_{2m}, \quad m \geq 2. \quad (\text{A.6})$$

A.2 Computing $\delta^2 R_A$

Taking a second variation of R_A gives

$$\delta^2 R_A = \text{Tr} (\delta \rho_A \delta [\log(\rho_A)]) + \text{Tr} (\rho_A \delta^2 [\log(\rho_A)]). \quad (\text{A.7})$$

⁸Iterating eq. (12) of [45] in such a way as to obtain all terms with two Y 's gives

$$\log(e^X e^Y) = X + Y + \sum_{k=1}^{\infty} R_k \int_0^1 dt \left(\text{Ad}_X + t \text{Ad}_Y + \sum_{m=1}^{\infty} R_m [\text{Ad}_X^{(m)} t \text{Ad}_Y] + \dots \right)^n Y \quad (\text{A.4})$$

where $R_n = (-1)^n B_n / n!$ and $\text{Ad}_Z W = [Z, W]$. After collecting all of the the terms that give nested commutators of the form $[Y X^{(n)} Y]$ we obtain the first equality in (A.5).

To evaluate the variation of the logarithm we first rewrite

$$\log(\rho_A + \delta\rho_A) = \log[e^{-H_A}(1 + e^{H_A}\delta\rho_A)] \quad (\text{A.8})$$

and apply (A.1) with $X = -H_A$ and $Y = \log(1 + e^{H_A}\delta\rho_A)$. We then Taylor expand Y using

$$\log(1 + Z) = Z - \frac{Z^2}{2} + O(Z^3). \quad (\text{A.9})$$

None of the $D_{n,k \geq 1}$ terms from (A.1) contribute to (A.7) because they can only appear in the second term of (A.7) and

$$\text{Tr} \left(\rho_A [H_A^{(k \geq 1)} (\rho_A^{-1} \delta\rho_A)^{(1)} H_A^{(n-k)} (\rho_A^{-1} \delta\rho_A)^{(1)}] \right) = \text{Tr} \left([H_A^{(k \geq 1)} \delta\rho_A^{(1)} H_A^{(n-k)} (\rho_A^{-1} \delta\rho_A)^{(1)}] \right), \quad (\text{A.10})$$

which vanishes by cyclicity of the trace. The remaining terms give

$$\begin{aligned} \delta^2 R_A &= \frac{1}{2} \text{Tr}(e^{H_A} \delta\rho_A^2) + \sum_{n=1}^{\infty} \left(\frac{B_n}{n!} \right) \text{Tr} \left(e^{H_A} [H_A^{(n)} \delta\rho_A^{(1)}] \delta\rho_A \right) \\ &\quad + \sum_{n=1}^{\infty} (-1)^n D_{n,0} \text{Tr} \left([(e^{H_A})^{(1)} H_A^{(n)} \delta\rho_A] \delta\rho_A \right). \end{aligned} \quad (\text{A.11})$$

Note that when $[H_A, \delta\rho_A] = 0$, (A.11) reduces to the term studied in [53].

We now evaluate the sums in (A.11). It is useful to rewrite the first trace as

$$\begin{aligned} \text{Tr} \left(e^{H_A} [H_A^{(n)} \delta\rho_A^{(1)}] \delta\rho_A \right) &= \sum_{j=0}^n (-1)^j \binom{n}{j} \text{Tr} \left(e^{H_A} H_A^{n-j} \delta\rho_A H_A^j \delta\rho_A \right) \\ &= \sum_{j=0}^n (-1)^j \binom{n}{j} \langle e^{H_A} H_A^{n-j} \delta\rho_A H_A^j \delta\rho_A e^{H_A} \rangle \\ &= \langle (e^{H_A} \delta\rho_A) (-K)^n (\delta\rho_A e^{H_A}) \rangle, \end{aligned} \quad (\text{A.12})$$

where the first line is straightforward to prove by induction. In the second equality we have inserted a factor of $e^{H_A} \rho_A = 1$ into the trace to convert it to an expectation value with respect to the reduced density matrix ρ_A . Since all of the operators inside the expectation value live on A , this is equivalent to the expectation value with respect to the full density matrix $\rho = |0\rangle\langle 0|$, which, as in the main text, is denoted by $\langle \dots \rangle$. The last equality follows from $H_A|0\rangle = H_{A^c}|0\rangle$, the definition $K := H_A - H_{A^c}$, and the binomial expansion.

Similarly, the second trace in (A.11) can be rewritten in the form

$$\begin{aligned} \text{Tr} \left([(e^{H_A})^{(1)} H_A^{(n)} \delta\rho_A] \delta\rho_A \right) &= \sum_{j=0}^n (-1)^j \binom{n}{j} \text{Tr} \left(H_A^{n-j} [e^{H_A}, \delta\rho_A] H_A^j \delta\rho_A \right) \\ &= \sum_{j=0}^n (-1)^j \binom{n}{j} \left(\langle (e^{H_A} \delta\rho_A) H_{A^c}^{n-j} H_A^j (\delta\rho_A e^{H_A}) \rangle - \langle (e^{H_A} \delta\rho_A) H_A^{n-j} H_{A^c}^j (\delta\rho_A e^{H_A}) \rangle \right) \\ &= [(-1)^n - 1] \langle (e^{H_A} \delta\rho_A) K^n (\delta\rho_A e^{H_A}) \rangle, \end{aligned} \quad (\text{A.13})$$

which clearly vanishes for even n . Inserting (A.12) and (A.13) into (A.11) and performing the sums gives

$$\delta^2 R_A = \langle e^{H_A} \delta \rho_A [f(K) - f'(K)] \delta \rho_A e^{H_A} \rangle, \quad (\text{A.14})$$

where f is the smooth function defined in (3.3). Note that here (and in previous expressions involving correlators) the operators H_A , $\delta \rho_A$ should be understood as tensor products with the identity operator on A^c . We computed the above sums using (A.3), (A.5) and the trick

$$\sum_{n=1}^{\infty} \frac{B_{n+1}}{n!} (-y)^n = -\partial_y \left(\sum_{n=0}^{\infty} \frac{B_n}{n!} (-y)^n - \frac{y}{2} \right) = -f'(y) + \frac{1}{2}. \quad (\text{A.15})$$

Our manipulations of the infinite sums (A.3) and (A.15) are somewhat formal since each sum converges only for $y \in (-2\pi, 2\pi)$ while the eigenvalues of H_A are unbounded. Nevertheless, since both the initial and final expressions involve only analytic functions of H_A , we expect that with appropriate care the final result can be justified for general H_A . The representation of $\log \rho$ used in e.g. appendix C of [5] may be useful for this purpose.

Note that since K is Hermitian and

$$f(y) - f'(y) = \frac{-1 + y + e^{-y}}{(1 - e^{-y})^2} > 0 \quad (\text{A.16})$$

for real y , the eigenvalues of $f(K) - f'(K)$ are real and positive. This fact implies $\delta^2 R_A \geq 0$, with equality if and only if $\delta \rho_A e^{H_A}$ annihilates the vacuum. But since the restriction of $|0\rangle$ to A is non-degenerate, this can occur only if $\delta \rho_A = 0$. Thus (A.14) is consistent with positivity of the relative entropy.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] T. Jacobson, *Thermodynamics of space-time: the Einstein equation of state*, *Phys. Rev. Lett.* **75** (1995) 1260 [[gr-qc/9504004](#)] [[INSPIRE](#)].
- [2] N. Lashkari, M.B. McDermott and M. Van Raamsdonk, *Gravitational dynamics from entanglement ‘thermodynamics’*, *JHEP* **04** (2014) 195 [[arXiv:1308.3716](#)] [[INSPIRE](#)].
- [3] T. Faulkner, M. Guica, T. Hartman, R.C. Myers and M. Van Raamsdonk, *Gravitation from entanglement in holographic CFTs*, *JHEP* **03** (2014) 051 [[arXiv:1312.7856](#)] [[INSPIRE](#)].
- [4] B. Swingle and M. Van Raamsdonk, *Universality of gravity from entanglement*, [arXiv:1405.2933](#) [[INSPIRE](#)].
- [5] T. Faulkner, *Bulk emergence and the RG flow of entanglement entropy*, *JHEP* **05** (2015) 033 [[arXiv:1412.5648](#)] [[INSPIRE](#)].
- [6] M. Nozaki, T. Numasawa, A. Prudenziati and T. Takayanagi, *Dynamics of entanglement entropy from Einstein equation*, *Phys. Rev. D* **88** (2013) 026012 [[arXiv:1304.7100](#)] [[INSPIRE](#)].

- [7] J. Bhattacharya and T. Takayanagi, *Entropic counterpart of perturbative Einstein equation*, *JHEP* **10** (2013) 219 [[arXiv:1308.3792](#)] [[INSPIRE](#)].
- [8] T. Jacobson, *Entanglement equilibrium and the Einstein equation*, [arXiv:1505.04753](#) [[INSPIRE](#)].
- [9] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [[hep-th/0603001](#)] [[INSPIRE](#)].
- [10] S. Ryu and T. Takayanagi, *Aspects of holographic entanglement entropy*, *JHEP* **08** (2006) 045 [[hep-th/0605073](#)] [[INSPIRE](#)].
- [11] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [[arXiv:0705.0016](#)] [[INSPIRE](#)].
- [12] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [[arXiv:1304.4926](#)] [[INSPIRE](#)].
- [13] F.M. Haehl, T. Hartman, D. Marolf, H. Maxfield and M. Rangamani, *Topological aspects of generalized gravitational entropy*, *JHEP* **05** (2015) 023 [[arXiv:1412.7561](#)] [[INSPIRE](#)].
- [14] S. Fischetti and D. Marolf, *Complex entangling surfaces for AdS and Lifshitz black holes?*, *Class. Quant. Grav.* **31** (2014) 214005 [[arXiv:1407.2900](#)] [[INSPIRE](#)].
- [15] S. Fischetti, D. Marolf and A.C. Wall, *A paucity of bulk entangling surfaces: AdS wormholes with de Sitter interiors*, *Class. Quant. Grav.* **32** (2015) 065011 [[arXiv:1409.6754](#)] [[INSPIRE](#)].
- [16] P.D. Hislop and R. Longo, *Modular structure of the local algebras associated with the free massless scalar field theory*, *Commun. Math. Phys.* **84** (1982) 71 [[INSPIRE](#)].
- [17] H. Casini, M. Huerta and R.C. Myers, *Towards a derivation of holographic entanglement entropy*, *JHEP* **05** (2011) 036 [[arXiv:1102.0440](#)] [[INSPIRE](#)].
- [18] M. Goykhman, *Entanglement entropy in 't Hooft model*, *Phys. Rev. D* **92** (2015) 025048 [[arXiv:1501.07590](#)] [[INSPIRE](#)].
- [19] D. Marolf, D. Minic and S.F. Ross, *Notes on space-time thermodynamics and the observer dependence of entropy*, *Phys. Rev. D* **69** (2004) 064006 [[hep-th/0310022](#)] [[INSPIRE](#)].
- [20] D. Marolf, *A few words on entropy, thermodynamics and horizons*, [hep-th/0410168](#) [[INSPIRE](#)].
- [21] D.D. Blanco, H. Casini, L.-Y. Hung and R.C. Myers, *Relative entropy and holography*, *JHEP* **08** (2013) 060 [[arXiv:1305.3182](#)] [[INSPIRE](#)].
- [22] J. Lin, M. Marcolli, H. Ooguri and B. Stoica, *Locality of gravitational systems from entanglement of conformal field theories*, *Phys. Rev. Lett.* **114** (2015) 221601 [[arXiv:1412.1879](#)] [[INSPIRE](#)].
- [23] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [[arXiv:1307.2892](#)] [[INSPIRE](#)].
- [24] I. Heemskerk, J. Penedones, J. Polchinski and J. Sully, *Holography from conformal field theory*, *JHEP* **10** (2009) 079 [[arXiv:0907.0151](#)] [[INSPIRE](#)].
- [25] T. Hartman, *Entanglement entropy at large central charge*, [arXiv:1303.6955](#) [[INSPIRE](#)].
- [26] T. Hartman, C.A. Keller and B. Stoica, *Universal spectrum of 2d conformal field theory in the large c limit*, *JHEP* **09** (2014) 118 [[arXiv:1405.5137](#)] [[INSPIRE](#)].

- [27] G. 't Hooft, *A planar diagram theory for strong interactions*, *Nucl. Phys. B* **72** (1974) 461 [[INSPIRE](#)].
- [28] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Local bulk operators in AdS/CFT: a boundary view of horizons and locality*, *Phys. Rev. D* **73** (2006) 086003 [[hep-th/0506118](#)] [[INSPIRE](#)].
- [29] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Holographic representation of local bulk operators*, *Phys. Rev. D* **74** (2006) 066009 [[hep-th/0606141](#)] [[INSPIRE](#)].
- [30] D. Kabat, G. Lifschytz and D.A. Lowe, *Constructing local bulk observables in interacting AdS/CFT*, *Phys. Rev. D* **83** (2011) 106009 [[arXiv:1102.2910](#)] [[INSPIRE](#)].
- [31] I. Heemskerk, D. Marolf, J. Polchinski and J. Sully, *Bulk and transhorizon measurements in AdS/CFT*, *JHEP* **10** (2012) 165 [[arXiv:1201.3664](#)] [[INSPIRE](#)].
- [32] D. Kabat and G. Lifschytz, *Decoding the hologram: scalar fields interacting with gravity*, *Phys. Rev. D* **89** (2014) 066010 [[arXiv:1311.3020](#)] [[INSPIRE](#)].
- [33] I.A. Morrison, *Boundary-to-bulk maps for AdS causal wedges and the Reeh-Schlieder property in holography*, *JHEP* **05** (2014) 053 [[arXiv:1403.3426](#)] [[INSPIRE](#)].
- [34] D. Kabat and G. Lifschytz, *Bulk equations of motion from CFT correlators*, *JHEP* **09** (2015) 059 [[arXiv:1505.03755](#)] [[INSPIRE](#)].
- [35] P.V. Buividovich and M.I. Polikarpov, *Entanglement entropy in gauge theories and the holographic principle for electric strings*, *Phys. Lett. B* **670** (2008) 141 [[arXiv:0806.3376](#)] [[INSPIRE](#)].
- [36] W. Donnelly, *Decomposition of entanglement entropy in lattice gauge theory*, *Phys. Rev. D* **85** (2012) 085004 [[arXiv:1109.0036](#)] [[INSPIRE](#)].
- [37] H. Casini, M. Huerta and J.A. Rosabal, *Remarks on entanglement entropy for gauge fields*, *Phys. Rev. D* **89** (2014) 085012 [[arXiv:1312.1183](#)] [[INSPIRE](#)].
- [38] H. Casini and M. Huerta, *Entanglement entropy for a Maxwell field: numerical calculation on a two dimensional lattice*, *Phys. Rev. D* **90** (2014) 105013 [[arXiv:1406.2991](#)] [[INSPIRE](#)].
- [39] W. Donnelly, *Entanglement entropy and nonabelian gauge symmetry*, *Class. Quant. Grav.* **31** (2014) 214003 [[arXiv:1406.7304](#)] [[INSPIRE](#)].
- [40] W. Donnelly and A.C. Wall, *Entanglement entropy of electromagnetic edge modes*, *Phys. Rev. Lett.* **114** (2015) 111603 [[arXiv:1412.1895](#)] [[INSPIRE](#)].
- [41] L.-Y. Hung and Y. Wan, *Revisiting entanglement entropy of lattice gauge theories*, *JHEP* **04** (2015) 122 [[arXiv:1501.04389](#)] [[INSPIRE](#)].
- [42] S. Ghosh, R.M. Soni and S.P. Trivedi, *On the entanglement entropy for gauge theories*, *JHEP* **09** (2015) 069 [[arXiv:1501.02593](#)] [[INSPIRE](#)].
- [43] S. Aoki et al., *On the definition of entanglement entropy in lattice gauge theories*, *JHEP* **06** (2015) 187 [[arXiv:1502.04267](#)] [[INSPIRE](#)].
- [44] W. Donnelly and A.C. Wall, *Geometric entropy and edge modes of the electromagnetic field*, [arXiv:1506.05792](#) [[INSPIRE](#)].
- [45] D. Gensing and G. Gensing, *Generalized Campbell-Baker-Hausdorff formula, path ordering and Bernoulli numbers*, *Z. Phys. C* **33** (1986) 307 [[INSPIRE](#)].

- [46] R. Haag, *Local quantum physics: fields, particles, algebras*, Springer, Berlin Germany (1992) [[INSPIRE](#)].
- [47] K. Papadodimas and S. Raju, *Black hole interior in the holographic correspondence and the information paradox*, *Phys. Rev. Lett.* **112** (2014) 051301 [[arXiv:1310.6334](#)] [[INSPIRE](#)].
- [48] W. Rossmann, *Lie groups: an introduction through linear groups*, Oxford University Press, Oxford U.K. (2006).
- [49] L.D. Landau, *The classical theory of fields*, Course of Theoretical Physics Series, volume 2, fourth edition, Pergamon Press (1987).
- [50] S. Hollands and R.M. Wald, *Stability of black holes and black branes*, *Commun. Math. Phys.* **321** (2013) 629 [[arXiv:1201.0463](#)] [[INSPIRE](#)].
- [51] N. Lashkari and M. Van Raamsdonk, *Canonical energy is quantum Fisher information*, [arXiv:1508.00897](#) [[INSPIRE](#)].
- [52] N. Engelhardt and A.C. Wall, *Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime*, *JHEP* **01** (2015) 073 [[arXiv:1408.3203](#)] [[INSPIRE](#)].
- [53] V. Rosenhaus and M. Smolkin, *Entanglement entropy: a perturbative calculation*, *JHEP* **12** (2014) 179 [[arXiv:1403.3733](#)] [[INSPIRE](#)].
- [54] V.E. Hubeny and M. Rangamani, *Causal holographic information*, *JHEP* **06** (2012) 114 [[arXiv:1204.1698](#)] [[INSPIRE](#)].
- [55] B. Freivogel and B. Mosk, *Properties of causal holographic information*, *JHEP* **09** (2013) 100 [[arXiv:1304.7229](#)] [[INSPIRE](#)].
- [56] W.R. Kelly and A.C. Wall, *Coarse-grained entropy and causal holographic information in AdS/CFT*, *JHEP* **03** (2014) 118 [[arXiv:1309.3610](#)] [[INSPIRE](#)].
- [57] K. Dilcher, *Sums of products of Bernoulli numbers*, *J. Number Theory* **60** (1996) 23.